

Dynamic Memory Networks

1 GRU

$$\mathbf{Z}_t = \sigma \left(\mathbf{X}_t \mathbf{W}^{(z)} + \mathbf{H}_{t-1} \mathbf{U}^{(z)} + (\mathbf{b}^{(z)})^T \right) \quad (1)$$

$$\mathbf{R}_t = \sigma \left(\mathbf{X}_t \mathbf{W}^{(r)} + \mathbf{H}_{t-1} \mathbf{U}^{(r)} + (\mathbf{b}^{(r)})^T \right) \quad (2)$$

$$\tilde{\mathbf{H}}_t = \tanh \left(\mathbf{X}_t \mathbf{W} + \mathbf{R} \odot \mathbf{H}_{t-1} \mathbf{U} + \mathbf{b}^T \right) \quad (3)$$

$$\mathbf{H}_t = \mathbf{Z}_t \odot \mathbf{H}_{t-1} + (1 - \mathbf{Z}_t) \odot \tilde{\mathbf{H}}_t \quad (4)$$

$$\mathbf{H}_0 = \mathbf{0} \quad (5)$$

Let $\mathbf{H}_t = \text{GRU}(\mathbf{X}_t, \mathbf{H}_{t-1})$ resume the previous equations.